Impact of Severe Weather Events on Public Health and Economy in the United States.

Synopsis

The U.S. National Oceanic and Atmospheric Administration's (NOAA) storm database tracks characteristics of major storms and weather events in the United States, include when and where they occur, as well as estimates of any fatalities, injuries and property damage. This report contains the exploratory analysis results on the health and economic impact by the severe weather events based on the data from NOAA database.

Required packages

```
library(R.utils)
library(ggplot2)
library(gridExtra)
library(knitr)
```

Loading data in R

Download and unzip the data in the working directory.

```
if (!"stormData.csv.bz2" %in% dir("./")) {
    download.file("https://d396qusza40orc.cloudfront.net
                  /repdata%2Fdata%2FStormData.csv.bz2"
                  , destfile = "stormData.csv.bz2")
    bunzip2("stormData.csv.bz2",overwrite=TRUE,remove=FALSE)
}
```

Load the data.

```
data<-read.csv("stormData.csv",sep=",")
head(data)
```

```
##   STATE__           BGN_DATE BGN_TIME TIME_ZONE COUNTY COUNTYNAME STATE
## 1       1  4/18/1950 0:00:00     0130       CST     97     MOBILE    AL
## 2       1  4/18/1950 0:00:00     0145       CST      3    BALDWIN    AL
## 3       1  2/20/1951 0:00:00     1600       CST     57    FAYETTE    AL
## 4       1   6/8/1951 0:00:00     0900       CST     89    MADISON    AL
## 5       1 11/15/1951 0:00:00     1500       CST     43    CULLMAN    AL
## 6       1 11/15/1951 0:00:00     2000       CST     77 LAUDERDALE    AL
##    EVTYPE BGN_RANGE BGN_AZI BGN_LOCATI END_DATE END_TIME COUNTY_END
## 1 TORNADO         0                                               0
## 2 TORNADO         0                                               0
## 3 TORNADO         0                                               0
## 4 TORNADO         0                                               0
## 5 TORNADO         0                                               0
## 6 TORNADO         0                                               0
##   COUNTYENDN END_RANGE END_AZI END_LOCATI LENGTH WIDTH F MAG FATALITIES
```

```
## 1         NA         0                        14.0   100 3    0          0
## 2         NA         0                         2.0   150 2    0          0
## 3         NA         0                         0.1   123 2    0          0
## 4         NA         0                         0.0   100 2    0          0
## 5         NA         0                         0.0   150 2    0          0
## 6         NA         0                         1.5   177 2    0          0
##   INJURIES PROPDMG PROPDMGEXP CROPDMG CROPDMGEXP WFO STATEOFFIC ZONENAMES
## 1       15    25.0          K       0
## 2        0     2.5          K       0
## 3        2    25.0          K       0
## 4        2     2.5          K       0
## 5        2     2.5          K       0
## 6        6     2.5          K       0
##   LATITUDE LONGITUDE LATITUDE_E LONGITUDE_ REMARKS REFNUM
## 1     3040      8812       3051       8806              1
## 2     3042      8755          0          0              2
## 3     3340      8742          0          0              3
## 4     3458      8626          0          0              4
## 5     3412      8642          0          0              5
## 6     3450      8748          0          0              6
```

Data Processing

Calcualte actual "PROPERTY_DAMAGE" and "CROP_DAMAGE" in billions US dollars and add them as columns in the dataset.

```r
levels(data$PROPDMGEXP)
```

```
##  [1] ""  "-" "?" "+" "0" "1" "2" "3" "4" "5" "6" "7" "8" "B" "h" "H" "K"
## [18] "m" "M"
```

```r
levels(data$PROPDMGEXP)<-c(1,0,0,0,1,10,100,1000,1e+04,1e+05,1e+06,1e+07,
                           1e+08,1e+09,1e+02,1e+02,1e+03,1e+06,1e+06)
levels(data$PROPDMGEXP)
```

```
##  [1] "1"     "0"     "10"    "100"   "1000"  "10000" "1e+05" "1e+06"
##  [9] "1e+07" "1e+08" "1e+09"
```

```r
data$PROPERTY_DAMAGE<-(data$PROPDMG*
                         as.numeric(as.character(data$PROPDMGEXP)))/1e+09
levels(data$CROPDMGEXP)
```

```
## [1] ""  "?" "0" "2" "B" "k" "K" "m" "M"
```

```r
levels(data$CROPDMGEXP)<-c(1,0,1,100,1e+09,1e+03,1e+03,1e+06,1e+06)
levels(data$CROPDMGEXP)
```

```
## [1] "1"     "0"     "100"   "1e+09" "1000"  "1e+06"
```

```
data$CROP_DAMAGE<-(data$CROPDMG*as.numeric(as.character(data$CROPDMGEXP)))/1e+09
```

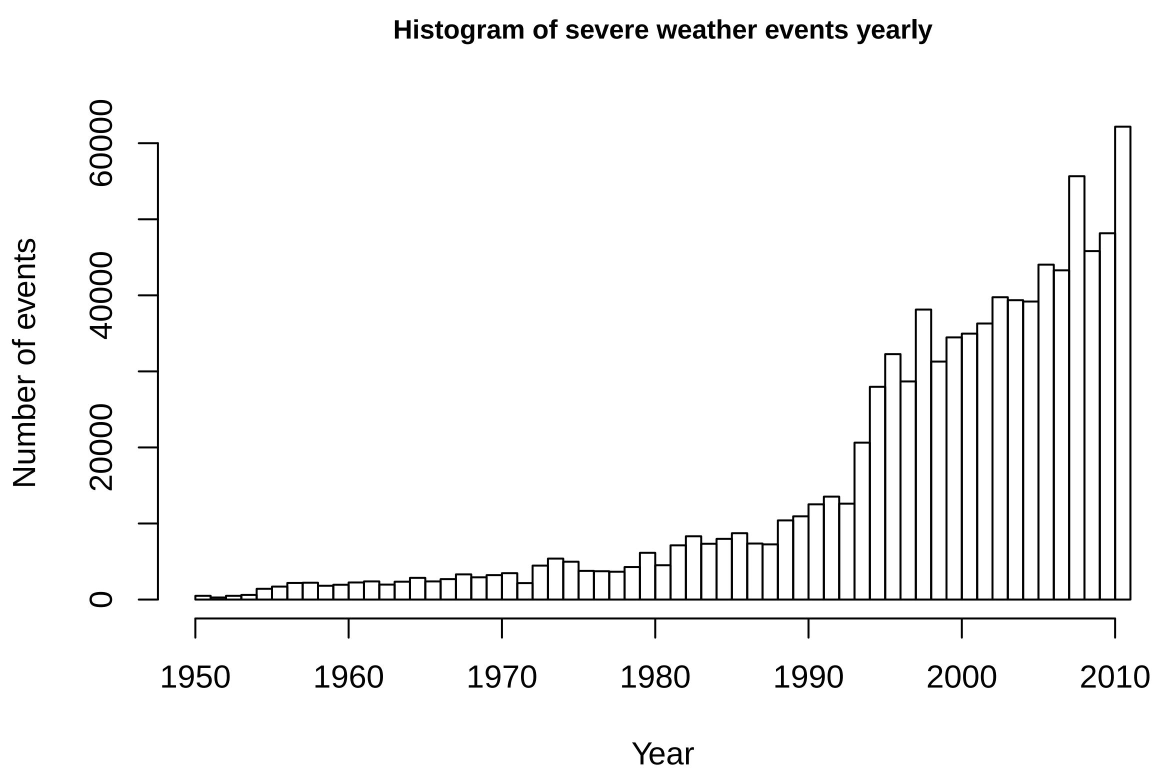
The events in the database start in the year 1950 and end in November 2011. In the earlier years of the database there are generally fewer events recorded, most likely due to a lack of good records. More recent years should be considered more complete.

Extract year from date and add a column "YEAR" in the data.

```
data$YEAR<-as.Date(as.character(data$BGN_DATE),format="%m/%d/%Y")
data$YEAR<-as.numeric(format(data$YEAR,"%Y"))
```

Make a histogram of years when physical disaster occured.

```
hist(data$YEAR,50,xlab="Year",main="Histogram of severe weather events yearly",
     cex.main=0.85,ylab="Number of events")
```



Based on the above histogram, we see that the number of events tracked starts to significantly increase around 1993-1994. So we use the subset of the data from 1993 to 2011 to get most out of good records.

Subset the data from 1993 to 2011.

```
tidyData<-subset(data,data$YEAR>=1993)
```

The names of the columns/variables

```r
names(tidyData)
```

```
##  [1] "STATE__"         "BGN_DATE"        "BGN_TIME"       
##  [4] "TIME_ZONE"       "COUNTY"          "COUNTYNAME"     
##  [7] "STATE"           "EVTYPE"          "BGN_RANGE"      
## [10] "BGN_AZI"         "BGN_LOCATI"      "END_DATE"       
## [13] "END_TIME"        "COUNTY_END"      "COUNTYENDN"     
## [16] "END_RANGE"       "END_AZI"         "END_LOCATI"     
## [19] "LENGTH"          "WIDTH"           "F"              
## [22] "MAG"             "FATALITIES"      "INJURIES"       
## [25] "PROPDMG"         "PROPDMGEXP"      "CROPDMG"        
## [28] "CROPDMGEXP"      "WFO"             "STATEOFFIC"     
## [31] "ZONENAMES"       "LATITUDE"        "LONGITUDE"      
## [34] "LATITUDE_E"      "LONGITUDE_"      "REMARKS"        
## [37] "REFNUM"          "PROPERTY_DAMAGE" "CROP_DAMAGE"    
## [40] "YEAR"
```

Results

Impact on Public Health

```r
fatalities<-aggregate(FATALITIES~EVTYPE,data=tidyData,FUN=sum)
fatalities<-fatalities[order(fatalities$FATALITIES,decreasing=TRUE),]
fatalities<-fatalities[1:10,]
row.names(fatalities)<-seq(1:10)
fatalities
```

```
##            EVTYPE FATALITIES
## 1  EXCESSIVE HEAT       1903
## 2         TORNADO       1621
## 3     FLASH FLOOD        978
## 4            HEAT        937
## 5       LIGHTNING        816
## 6           FLOOD        470
## 7     RIP CURRENT        368
## 8       HIGH WIND        248
## 9       TSTM WIND        241
## 10      AVALANCHE        224
```

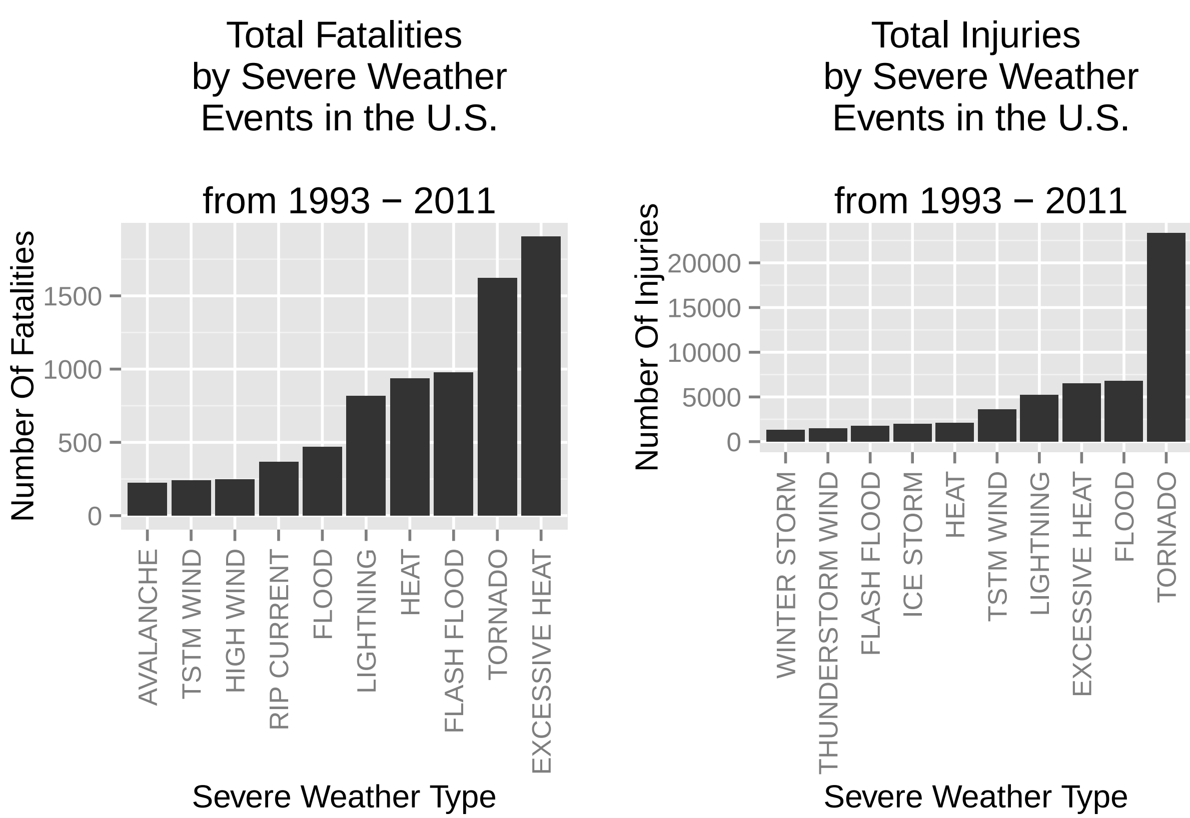
```r
injuries<-aggregate(INJURIES~EVTYPE,data=tidyData,FUN=sum)
injuries<-injuries[order(injuries$INJURIES,decreasing=TRUE),]
injuries<-injuries[1:10,]
row.names(injuries)<-seq(1:10)
injuries
```

```
##               EVTYPE INJURIES
## 1            TORNADO    23310
## 2              FLOOD     6789
## 3     EXCESSIVE HEAT     6525
## 4          LIGHTNING     5230
## 5          TSTM WIND     3631
```

```
## 6                HEAT     2100
## 7           ICE STORM     1975
## 8         FLASH FLOOD     1777
## 9   THUNDERSTORM WIND     1488
## 10       WINTER STORM     1321
```

```r
plotFatalities<-ggplot(fatalities,aes(reorder(EVTYPE,FATALITIES),FATALITIES))+
  geom_bar(stat="identity")+
  theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5))+
  labs(x="Severe Weather Type",y="Number Of Fatalities",
       title="Total Fatalities\n by Severe Weather\n Events in the U.S.
       \n from 1993 - 2011")
plotInjuries<-ggplot(injuries,aes(reorder(EVTYPE,INJURIES),INJURIES))+
  geom_bar(stat="identity")+
  theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5))+
  labs(x="Severe Weather Type",y="Number Of Injuries",
       title="Total Injuries\n by Severe Weather\n Events in the U.S.
       \n from 1993 - 2011")

grid.arrange(plotFatalities, plotInjuries, ncol = 2)
```



Based on the above Barplopts, we find that excessive heat and tornado cause most fatalities. Tornado causes most injuries in the United States from 1993 to 2011.

Impact on Economy

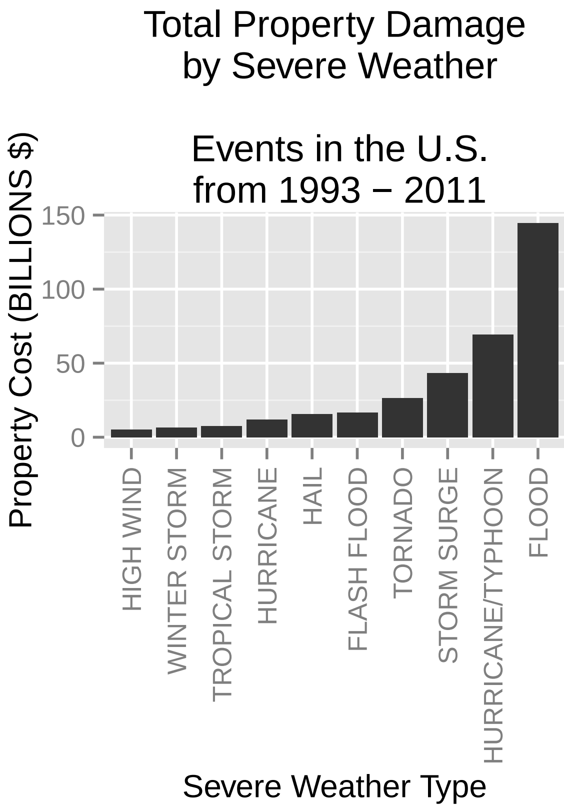
```r
property_DMG<-aggregate(PROPERTY_DAMAGE~EVTYPE,data=tidyData,FUN=sum)
property_DMG<-property_DMG[order(property_DMG$PROPERTY_DAMAGE,decreasing=TRUE),]
property_DMG<-property_DMG[1:10,]
row.names(property_DMG)<-seq(1:10)
names(property_DMG)[2]<-"PROPERTY_DAMAGE(billions $)"
property_DMG
```

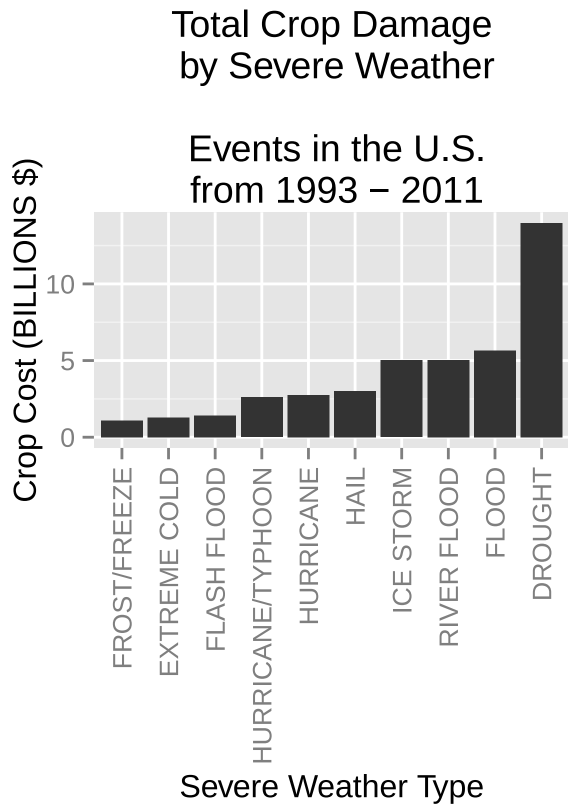
```
##                EVTYPE PROPERTY_DAMAGE(billions $)
## 1               FLOOD                  144.657710
## 2  HURRICANE/TYPHOON                   69.305840
## 3         STORM SURGE                   43.323536
## 4             TORNADO                   26.349182
## 5         FLASH FLOOD                   16.822674
## 6                HAIL                   15.735268
## 7           HURRICANE                   11.868319
## 8      TROPICAL STORM                    7.703891
## 9        WINTER STORM                    6.688497
## 10          HIGH WIND                    5.270046
```

```r
crop_DMG<-aggregate(CROP_DAMAGE~EVTYPE,data=tidyData,FUN=sum)
crop_DMG<-crop_DMG[order(crop_DMG$CROP_DAMAGE,decreasing=TRUE),]
crop_DMG<-crop_DMG[1:10,]
row.names(crop_DMG)<-seq(1:10)
names(crop_DMG)[2]<-"CROP_DAMAGE(billions $)"
crop_DMG
```

```
##                EVTYPE CROP_DAMAGE(billions $)
## 1             DROUGHT               13.972566
## 2               FLOOD                5.661968
## 3         RIVER FLOOD                5.029459
## 4           ICE STORM                5.022113
## 5                HAIL                3.025954
## 6           HURRICANE                2.741910
## 7  HURRICANE/TYPHOON                 2.607873
## 8         FLASH FLOOD                1.421317
## 9        EXTREME COLD                1.292973
## 10       FROST/FREEZE                1.094086
```

```r
plotProperty<-ggplot(property_DMG,aes(reorder(EVTYPE,property_DMG[,2]),
                                      property_DMG[,2]))+
  geom_bar(stat="identity")+
  theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5))+
  labs(x="Severe Weather Type",y="Property Cost (BILLIONS $)",
       title="Total Property Damage\n by Severe Weather
       \n Events in the U.S.\n from 1993 - 2011")
plotCrop<-ggplot(crop_DMG,aes(reorder(EVTYPE,crop_DMG[,2]),crop_DMG[,2]))+
  geom_bar(stat="identity")+
  theme(axis.text.x=element_text(angle=90,hjust=1,vjust=0.5))+
  labs(x="Severe Weather Type",y="Crop Cost (BILLIONS $)",
       title="Total Crop Damage\n by Severe Weather
       \n Events in the U.S.\n from 1993 - 2011")
grid.arrange(plotProperty, plotCrop, ncol = 2)
```





Based on the above barplots,across the United States, flood and typhoon have caused the greatest damage to properties. Drought and flood have caused the greatest damage to crops.